Key issues in reusing data

Patty Doran
UK Data Service
patty.doran@manchester.ac.uk
@Patty_NZ
Overview

- Data available for reuse
- Issues with reusing data
- Quantitative
  - Case study
  - What’s available
  - Key issues
- Qualitative
  - Case study
  - What’s available
  - Key issues
- Talk for ~35 min
- Questions
Reusing data

<table>
<thead>
<tr>
<th>Data collector</th>
<th>Secondary Analyst</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary research</td>
<td>Secondary research</td>
</tr>
<tr>
<td>Data collection &amp; analysis for planned purpose</td>
<td>Re-analysis (can be for a different purpose)</td>
</tr>
</tbody>
</table>

Data Sharing
<table>
<thead>
<tr>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>datasets impossible to create</td>
<td>do not have insider understanding of data and data collection</td>
</tr>
<tr>
<td>cost effective</td>
<td>effort to get to know the data</td>
</tr>
<tr>
<td>ethical issues about data collection dealt with</td>
<td>ethical issues about data use still apply (limit access to sensitive data)</td>
</tr>
<tr>
<td>do not need to re-contact data subjects</td>
<td>data may not match research question</td>
</tr>
<tr>
<td>reuse data used by others to make claims</td>
<td>cannot extend studies</td>
</tr>
</tbody>
</table>

Researchers reusing data need to:
1. Make effort to understand the data
2. Be pragmatic about whether data are good enough for your purpose
Research plan

Research Question

Locate data

Evaluate data

Analyse
To understand data and its suitability for your research you must understand:

- what information was collected
- from whom
- when and where
- any changes to the raw data before being archived

Use documentation:
- user guides
- questionnaires
- interview schedules
Quantitative data

- Case study
- What data is available?
- What sorts of documentation are available?
- Key issues
Case Study: Crime Survey for England and Wales (formerly the British Crime Survey)

- Example of a repeated cross-sectional survey
- Important source of information about crime (Crime statistics independent from police records)
- Annual survey (interviews throughout the year)
- $n \approx 35,000$ individuals aged 16+ and 3,000 aged 10-15
- Victim of a crime in the previous 12 months
- Questions covering demographics and crime-related subjects such as attitudes to the police and the criminal justice system (CJS)
### Different levels of access

- **A standard End User Licence (download after registering)**
- **Secure Access (for sensitive information)**

#### Case Study: Crime Survey for England and Wales

- Data stored as individual anonymised records

#### Data Table

<table>
<thead>
<tr>
<th>rowlabel</th>
<th>split</th>
<th>sex</th>
<th>yrsarea</th>
<th>work2</th>
<th>tenure1</th>
<th>livharm1</th>
<th>agegrp7</th>
<th>ethgrp2a</th>
<th>educat3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>137068050 A</td>
<td>Female</td>
<td>20 years</td>
<td>Yes</td>
<td>Buying it</td>
<td>Single</td>
<td>45–54</td>
<td>White</td>
<td>Degree or...</td>
</tr>
<tr>
<td>2</td>
<td>147461190 C</td>
<td>Female</td>
<td>10 years</td>
<td>No</td>
<td>Own it out</td>
<td>Married/C...</td>
<td>55–64</td>
<td>White</td>
<td>Degree or...</td>
</tr>
<tr>
<td>3</td>
<td>137116250 A</td>
<td>Female</td>
<td>20 years</td>
<td>No</td>
<td>Rent it</td>
<td>Widowed</td>
<td>55–64</td>
<td>White</td>
<td>Degree or...</td>
</tr>
<tr>
<td>4</td>
<td>147354190 C</td>
<td>Female</td>
<td>20 years</td>
<td>Yes</td>
<td>Buying it</td>
<td>Married/C...</td>
<td>55–64</td>
<td>White</td>
<td>Degree or...</td>
</tr>
<tr>
<td>5</td>
<td>137061230 C</td>
<td>Female</td>
<td>20 years</td>
<td>No</td>
<td>Rent it</td>
<td>Widowed</td>
<td>65–74</td>
<td>White</td>
<td>None</td>
</tr>
<tr>
<td>6</td>
<td>13689230 C</td>
<td>Female</td>
<td>20 years</td>
<td>No</td>
<td>Own it out</td>
<td>Married/C...</td>
<td>65–74</td>
<td>White</td>
<td>O level/G...</td>
</tr>
<tr>
<td>7</td>
<td>135507330 A</td>
<td>Male</td>
<td>10 years</td>
<td>Yes</td>
<td>Rent it</td>
<td>Married/C...</td>
<td>45–54</td>
<td>White</td>
<td>None</td>
</tr>
<tr>
<td>8</td>
<td>136450220 B</td>
<td>Female</td>
<td>5 years</td>
<td>Yes</td>
<td>Own it out</td>
<td>Married/C...</td>
<td>55–64</td>
<td>White</td>
<td>Degree or...</td>
</tr>
<tr>
<td>9</td>
<td>136111200 D</td>
<td>Male</td>
<td>20 years</td>
<td>No</td>
<td>Own it out</td>
<td>Married/C...</td>
<td>55–64</td>
<td>White</td>
<td>Degree or...</td>
</tr>
<tr>
<td>10</td>
<td>136599250 A</td>
<td>Male</td>
<td>20 years</td>
<td>No</td>
<td>Own it out</td>
<td>Married/C...</td>
<td>75+</td>
<td>White</td>
<td>Apprentice...</td>
</tr>
<tr>
<td>11</td>
<td>136229130 A</td>
<td>Male</td>
<td>3 years</td>
<td>Yes</td>
<td>Buying it</td>
<td>Single</td>
<td>25–34</td>
<td>White</td>
<td>Degree or...</td>
</tr>
<tr>
<td>12</td>
<td>136947260 B</td>
<td>Male</td>
<td>5 years</td>
<td>Yes</td>
<td>Own it out</td>
<td>Married/C...</td>
<td>75+</td>
<td>White</td>
<td>Apprentice...</td>
</tr>
</tbody>
</table>


**British Crime Survey 2009/10**
- introduction of disability measures
- (a special licence version of the data)
- n = 46,398 adults 16+
- 9,037 had at least one limiting disability

**Findings:**

- Adjusting for age, sex, socio-economic characteristics...
  - disability increases risk of experiencing violence
  - levels of victimisation highest amongst those with mental health problems
  - estimated 116,000 victims of violence attributable to disability

http://www.plosone.org/article/info:doi/10.1371/journal.pone.0055952
## Quantitative data by type of analysis

<table>
<thead>
<tr>
<th>Type of analysis</th>
<th>Type of data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individuals, families, households, businesses: one point in time</td>
<td>Cross-sectional surveys (e.g. Time Use Survey) and census microdata</td>
</tr>
<tr>
<td>Individuals, families, households, businesses: multiple points in time</td>
<td>Repeated cross-sections (e.g. different years of Crime Survey for England and Wales, Labour Force Survey)</td>
</tr>
<tr>
<td>Individuals, families, households: over time</td>
<td>Longitudinal data (e.g. Understanding Society)</td>
</tr>
<tr>
<td>Small geographic areas</td>
<td>Census aggregate data, flow data</td>
</tr>
<tr>
<td>Comparing countries (over time)</td>
<td>International time series (e.g. World Bank Indicators)</td>
</tr>
</tbody>
</table>
Where, when and why?

• Start with catalogue record

Date & Geographic coverage is in the catalogue record for most studies

Access to documentation including user guide or technical report
Who was asked what…

- Computer Aided Interviewing (CAI) makes it easy to send respondents through the questionnaire by different routes
- Many questions may only be applicable to some of the sample
And what was done with data afterwards?

- derived variables are created from the ‘raw’ data
Sampling considerations

- Surveys are based on samples
- Is the sample representative?
  - Who is included (adults, those in private address)?
  - Response rate and bias
  - Do you need to apply a weight to make the data representative?
- Do I have enough cases to make a precise estimate?
  - Important for small sub-populations

For more information, see UK Data Service guides to Survey weights and Complex sampling
Cite the Data

- Citations provided for every record and dataset
- Use the UK Data Service citation tool to copy and paste the correct citation

British Social Attitudes Survey, 2013

Title: British Social Attitudes Survey, 2013
Alternative title: BSA
Study number (SN): 7500
Access: These data are safeguarded
Persistent identifier (DOI): 10.5255/UKDA-SN-7500-1
Series: British Social Attitudes Survey
Principal investigator(s): NatCen Social Research

The citation for this study is:
References and resources

General discussions:


UK Data Service resources:

• Data Skills Modules: https://www.ukdataservice.ac.uk/use-data/data-skills-modules
• Secondary analysis web pages: https://www.ukdataservice.ac.uk/use-data/secondary-analysis
• Using Survey Data: https://www.ukdataservice.ac.uk/media/455510/using-survey-data.pdf
• What is weighting?: http://ukdataservice.ac.uk/media/285227/weighting.pdf
• What is complex sample design: http://ukdataservice.ac.uk/media/398743/complexsampleguide.pdf
Reusing qualitative data

• Types of qualitative data reuse

• Case study of reusing UK Data Service data: School Leavers Study

• Getting started with reusing qualitative data
  • Key issue: documentation and sampling
  • Key issue: searching

• Finding data – browse, search, QualiBank
Many ways of re-using data

• Description
  - Literature review with data
  - See more data, not limited to published extracts
• Research design and methodological advancement
  - Study sampling, data collection methods, topics guides
  - Compare original interview schedule questions actually asked in the interview
• Reanalysis—a research question different to the original research
  - Very different – comparative keyword analysis of illness narratives
    Seale and Charteris-Black (2008)
  - Similar topics, but with different focus (Bornat et al. 2012)
• Learning and Teaching (Haynes and Jones 2012)
SN 4867: School Leavers Study

Principal investigator: Ray Pahl

In 1978, teachers at a comprehensive school on the Isle of Sheppey were asked to set an essay about 10 days before pupils were due to leave school. The essay asked students to imagine that they were nearing the end of their life, and that something had made them think back to the time when they left school. They were then asked to write an imaginary account of their life over the next 30 or 40 years.
School Leavers Re-Study: Imagining the Future

Re-users: Graham Crow and Dawn Lyon, University of Kent and University of Southampton
I want you to imagine that you are towards the end of your life. Look back over your life and say what happened to you. Don't write a very exaggerated story, just tell the straightforward story of your life as it might really be. Of course you cannot know what is going to happen to you, but you can describe the sort of thing that could happen if things go as you expect or hope. Spread your story over your whole life from the time of leaving school. Continue on another sheet as necessary.

My Past

I was sitting in front of the fire smoking my pipe, just thinking of the past. I was nearly seventeen when I left school. I had an apprenticeship in the Tudor Glass, an electrical apprenticeship lasting four years. I served my four years and got my diploma for an electrical engineer. I was now twenty and courting a dearly lovely girl. I had to work in London in my next job and came back home every weekend. Jobs were scarce so you had to take what you can get. I didn’t want to go to London, I was actually and honestly a country man. I was a country man. I was a country man.
Living and Working on Sheppey

1978: **141 essays** (89 boys and 52 girls)

2009-10: The exercise was repeated by the Living and Working on Sheppey project and **110 essays** (55 boys and 55 girls) were gathered from school pupils and members of youth groups on the Isle of Sheppey to compare to the earlier ones.

Together, these two sets of essays shed light on the aspirations of Sheppey’s young people (and young people more generally) and cover a range of topics including
• Health, Education, Career, Family and Leisure
<table>
<thead>
<tr>
<th>1978</th>
<th>2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mundane &amp; grounded jobs; gradual career progression; periods of unemployment.</td>
<td>Well-paid, instantaneous jobs; choice but uncertainty; influence of celebrity culture.</td>
</tr>
<tr>
<td>“It was hard finding a job, I failed a few chances, but eventually got what I wanted locally, a craft apprenticeship.” (Essay 27, male)</td>
<td>“I could not decide what to do. Then it came across my mind that I should be a body builder, and be good enough to enter the Mr Olympia contest. After that I went to an athletics club, within a few months I was scouted for the Olympics for the 200 metre sprint.” (Essay 10, male)</td>
</tr>
<tr>
<td>“I was on the dole for six months after leaving school, until I got a job in a garage.” (Essay 42, male)</td>
<td>“I was 20 now living the dream I had a amazing band...I had toured the world 3 times sold 4 million records.” (Essay 30, male)</td>
</tr>
<tr>
<td>“When we found a house it was a semi-detached in Sittingbourne. I wanted to live in Italy but that was asking for too much.” (Essay 56, male)</td>
<td>“I arrive at my 3-bedroom luxury villa; I land my helicopter on my own heli-pad and walk inside. I grab my keys and jump in my Bentley Continental GTS.” (Essay 40, male)</td>
</tr>
<tr>
<td>“I longed for something exciting and challenging. But yet again I had to settle for second best. I began working in a large clothes factory.” (Essay 104, female)</td>
<td>“In my future I want to become either: a dance teacher, hairdresser, or a Professional Show Jumper/horse rider. If I do become a dancer my dream would be to dance for Beyoncé or someone really famous.” (Essay 61, female)</td>
</tr>
</tbody>
</table>
Living and Working on Sheppey

http://www.livingandworkingonsheppey.co.uk/
School Leavers Study: publications


Getting started

• Orient yourself to the original research project
  - Documentation and metadata

• Understand the structure of the original data
  - Contexts – at 3 levels
  - Sampling and recruitment

• Become familiar with the project data as a whole
  - Logic of subsampling if volume prohibits reading it all
  - Both inductive and deductive strategies can work
Publications and other resources for reusing data

http://ukdataservice.ac.uk/use-data/secondary-analysis/reusing-qualitative-data.aspx
Find data: Key data

Find data: Search data catalogue
QualiBank

Search and browse a selection of qualitative surveys, interview and open-ended questions.

Results per page: 10
Displaying 1-10 of 146 results

   SN8226 Pioneers, 1996-2012
   ...Web Resource: Book Reference for “Poverty in the United Kingdom: A Survey of Household Resources and Standards of Living”...
   Access this collection from Discover

2. Summary of the interview with Peter Townsend
   SN8226 Pioneers, 1996-2012
   ... work, because of the need to please both left and right wing; the board included both a Labour MP and a Conservative MP, Angus Maude. This was an important early experience of the need to fight against the bureaucratisation of social science. First PEP pamphlet he did was Poverty: Ten Years After...
   Access this collection from Discover

3. Essay by 5
   SN4867 School Leavers Study, 1978
   ... and a permanent permanent home on one side was the trap of depression depression and poverty poverty and on the other a high chance of not having a family I chose to stay as I was a hoop and if the opportunity opportunity came to have a job I liked where I could still go out and dig but still have...
   Access this collection from Discover
I got a job as assistant site manager the pay wasn’t much but I was really happy the odd girlfriend came and went but as I spent a great deal of time in a tent or digging holes I didn’t get too attached but one day I made the mistake and decided to settle and have a home so I tried to get a post somewhere in England I had been fascinated with the past and since basics and history I narrowly scraped passes and by a prolonged stay away from home but I wasn’t to be common with this how my friendships have been

The archaeological site who been in a similar kind of position we had very similar tastes in most things and we both shared nearly the same views of life.
Cite the Data

A unique citation reference has been generated based on your selection.


Select the text above

You can copy and paste this citation as required in your outputs. This citation includes a URL which will link directly back to this page, where the cited text will be highlighted.

Show preview of citation URL in action
Key issues in reusing data - summary

• Many benefits
  - large scale research, ethics completed, cost effective

• Iterative
  - research questions, locate and evaluate data, analysis

• Context
  - consider methodology and prior research

• Plan and prepare
  - understand data through study documentation

• Support
  - more information available in the UK Data Service website

• Cite the data
Forthcoming Events

• **Webinars**
  - 26 May, 11.30-12.30pm: Dissertation projects: Introduction to secondary analysis for qualitative and quantitative data
  - 28 May, 3-4pm: Data management basics
  - 2 June, 1-2pm: Introduction to Text-Mining

• **Online Conferences**
  - 30 June: Health Studies User Conference
  - 8 July: Family Finance Surveys User Conference

• **UK Data Service Dissertation Award**
  - Deadline: 3 June
  - The Award recognises outstanding undergraduate student dissertations that include excellent reuse of data from the UK Data Service.
Questions

Website > Use Data > Secondary Analysis
https://www.ukdataservice.ac.uk/use-data/secondary-analysis.aspx

e-mail
help@ukdataservice.ac.uk

Video guides and webinars:
https://www.youtube.com/user/UKDATASERVICE

To follow UK Data Service on Twitter:
@UKDataService